STAT 201B: Probability

Feynman Liang*
Department of Statistics, UC Berkeley

Last updated: January 30, 2020

Contents

1 Lecture 3: More on Markov chains

2020-01-28

A few more things about Markov chains which we can do with little cost, with general state space $(S, \mathcal{S})$, discrete time $n = 0, 1, \ldots$.

General view of how to construct a sequence of RVs $(X_i)_{i \geq 0}$:

1. Make $X_0 \sim \lambda$, $\Pr(X_0 \in A) = \lambda(A)$ where $\lambda = \delta_x =$ Dirac unit mass at x.
2. Given $X_0 = x$, we have a Markov transition kernel

$$P(x_0, \cdot) = \Pr(X_1 \in \cdot \mid X_0 = x_0)$$

Construct (X_0, X_1) using Fubini

$$\mathbb{E} g(X_0, X_1) = \int_S \lambda(dx_0) \int_S P(x_0, dx_1) g(x_0, x_1)$$

3. Given $X_0 = x_0$ and $X_1 = x_1$, make X_2 according to

$$P(x_0, x_1, \cdot) = \Pr(X_2 \in \cdot \mid X_0 = x_0, X_1 = x_1)$$

We can do that! Nothing new required since

$$X_0 \mapsto (X_0, X_1), \qquad \lambda \mapsto \lambda \otimes P \text{ on } S \times S$$

Now

$$\mathbb{E} g(X_0, X_1, X_2) = \underbrace{\int_S \lambda(dx_0) \int_S P(x_0, dx_1)}_{\int_{S \times S} (\lambda \otimes P)(dx_0, dx_1)} \int_S P(x_0, x_1, dx_2) g(x_0, x_1, x_2)$$

if S is nice, every stochastic process $(X_i)_{i \geq 0}$ is "like" (i.e. FDDs are equal in distribution) the above construction.

Definition 1

Such a process is called *Markov* iff it is equal in distribution to a process made with a kernel

$$P(dx_{n+1} \mid x_0, \ldots, x_n) = P_n(dx_{n+1} \mid x_n)$$

If $P_n \equiv P$, then the process is called *homogeneous*.

*feynman@berkeley.edu

- Not mmuch general theory of inhomogeneous
- We will always henceforth assume homogeneous
- Mostly, discrete S
- Techniques for general S
  - Figure it out for finite/countable S
  - Write the args so they wrook more generally
- Note: if you upgrade $X_n = Y_n := (X_0, \dots, X_n)$
- Technical point: can only use Fubini to prescribe FDDs. From our schemme, we create consistent FDDs and are left with two options:
  1. Make the chain explicitly as $X_n =$ some measurable function on $(\Omega, \mathcal{F}, \Pr)$. For example, for IID uniforms take $X_n = F^{-1}(U_n)$ for an IID sequence U_n. Then you have RVs $(X_i)_{i\geq 0}$ on $(\Omega, \mathcal{F}, \Pr)$ which implies a law on $\prod_{i\geq 0} S$ given by
  $$\omega \mapsto (X_0(\omega), X_1(\omega), \dots)$$
  2. If not, we need Kolmogorov consistency or better (Ionescu-Tulcea Theorem, which says the infinite process exists with no regularity on $(S, \mathcal{S})$).

Fix $(S, \mathcal{S})$ and Markov kernel P. For $f \geq 0$ bounded measurable

$$(Pf)(x) := \int_S P(x, \cdot) f(\cdot) = \mathbb{E}[f(X_1) \mid X_0 = x]$$

Convention: We write $\mathbb{E}[f(X_1) \mid X_0 = x] = \mathbb{E}_x f(X_1)$, or more generally $\mathbb{E}[f(X_1) \mid X_0 \sim \lambda] = \mathbb{E}_\lambda f(X_1)$.

Example 2

$\mathbb{E}_\lambda f(X_0, X_1, X_2) = \int_S \lambda(dx) \mathbb{E}_x f(x, X_1, X_2)$ by the Fubini theorem. This is an example with a regular conditional distribution (RCD) for $(X_0, X_1, X_2 \mid X_0 = x)$.

More generally, anytime we see an expectation involving an integral we are really just doing conditioning. Most things in Markov chains are obvious by writing down Fubini.

A special case: $\mathbb{E}_x f(X_1) = f(x)$ for all $x \in S$ means that $(P - I)f = 0$ which means that $(f(X_0), f(X_1))$ is a martingale pair relative to $\sigma(X_0)$, $\sigma(X_0, X_1)$. More generally, for any initial distribution λ the collection $(f(X_i))_{i\geq 0}$ is a martingale relative to $\mathcal{F}_n = \sigma(X_i : i \leq n)$. For a martingale

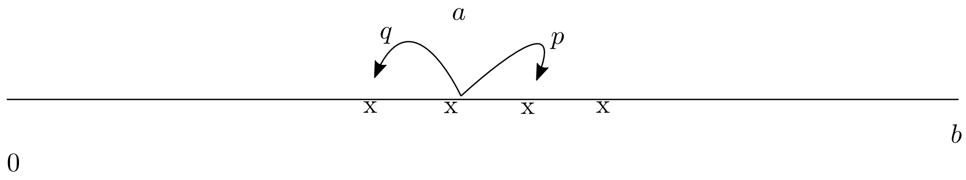
$$\mathbb{E}[f(X_{n+1} \mid X_0, \dots, X_n] = g(X_n)$$

by the time homogeneous markov property, we must have $g = Pf$. So if $Pf = f$, then $\mathbb{E}[\cdot] = f(X_n)$. This is the definition of a MG.

Example 3 (*Gambler's Ruin*)

{eg:gamblers-ruin}

$0 \leq a \leq b$, Gambler starts with a dollars and at each turn moves $+1$ with probability p and -1 with probability q. Continue until reach 0 or b and then stop.



{fig:gamblers-ruin}

Figure 1: Gambler's ruin

We can write this as $S_0 = a$ and $S_n = a + \sum_i X_i$ where X_i are ± 1 with probabilities p/q. Define (the stopping time)

$$T_{0b} = \min\{n \geq 1 : S_n \in \{0, b\}\}$$

Our MC is then $(S_{n \wedge T_{0b}} : n \geq 0)$. Viewed as a MC, we have an absorbing chain and are interested in $\Pr_a[\text{absorbed at } b]$.

Alternatively, viewed as a RW we are interested in

$$\Pr_a[S_{T_{0b}} = b]$$

By Wald's identity

$$\mathbb{E}[S_{T_{0b}} - a] = \mathbb{E}[X]\mathbb{E}[T_{0b}]$$

For fair case, $p = q = 1/2$ and $\mathbb{E}X = 0$.

$$(b-a)\Pr_a[\text{absorb at b}] + (0-a)\Pr_a[\text{absorb at 0}] + \underbrace{??}_{0} = 0$$

$$\Pr_a[\text{absorb at b}] + \Pr_a[\text{absorb at 0}] = 1$$

How do we know it eventually hits and doesn't oscillate forever? Bounded by Geometric RV, so goiing to be finite with probability 1 (could have use martingale convergence, but dont need).

In the unfair case, we need a better martingale (Wald MG). Try an exponential martingale

$$Z^{S_n} \text{ for suitable } Z$$

To find Z, we would like

$$\begin{aligned}
\mathbb{E}[Z^{\overbrace{=S_n + X_{n+1}}^{S_{n+1}}} \mid S_0, \ldots, S_n] &= Z^{S_n} \\
\Leftrightarrow \mathbb{E}Z^{X_{n+1}} &= 1 \\
\Rightarrow Z &= q/p
\end{aligned}$$

Applying Wald's identity again, we solve the system

$$\begin{aligned}
\left(\frac{q}{p}\right)^a &= \Pr_a[\text{abs at b}](q/p)^b + \Pr_a[\text{abs at 0}](q/p)^0 \\
1 &= \Pr_a[\cdot] + \Pr_a[\cdot]
\end{aligned}$$

Within the Markov chain frammework, we say that we seek a harmonic function h such that $h = Ph$ and $h(0) = 0$, $h(b) = 1$. Then

$$\Pr_a[\text{hit b before 0}] = ??$$

Notice $h_0(x) = (q/p)^x$ and $h_1(x) \equiv 1$ both solve $h = Ph$, so finding the linear combination that satisfies boundary conditions yields

$$h(x) = \frac{(q/p)^x - (q/p)^0}{(q/p)^b - (q/p)^0}$$

Theorem 4

Suppose you have a MC with absorbing states (i.e. boundary states) $B \subset S$ and target states $A \subset B$.

$$\Pr_a [X_n \in A \text{ for all large } n]$$

Let

$$h_A(x) = \Pr_x[\text{absorved in } A]$$

Argue $h(x) = h_A(x)$ must solve $h = Ph$ using a one-step analysis:

$$h(x) = \sum_{y \in S} P(x, y)h(y)$$

by conditioning on $X_1 = y$

Notice $h_A(x) = \lim_n P^n \mathbb{1}_A$ and

$$(P^n \mathbb{1}_A)(x) = \Pr_x[X_n \in A]$$

Since A is absorbing, $X_n \in A \Rightarrow X_{n+1} \in A$ so $\{X_n \in A\}$ is increasing, so $h_A = Ph_A$.

Also, $h_A = \mathbb{1}_A$ on B because B is absorbing.

Since Dirichlet BVP has unique soution, we hope/expect that there is a unique solution of $h = Ph$, $h = \mathbb{1}_A$ on B, and that in this case $h = h_A$.

2 Lecture 4

2020-01-30

Proposition 5 (*Simple Markov property*)

For any fixed time n

$$(X_n, X_{n+1}, \ldots \mid X_n = x) \stackrel{d}{=} (X_0, X_1, \ldots \mid X_0 = x)$$

Proof. To show $X \stackrel{d}{=} Y$, show $\Pr(X \in A) = \Pr(Y \in A)$ for all A in a π-system. Here π-system is FDDs, so (by Kolmogorov) its enough to take

$$A = (X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m)$$

for any choice of $x_0, \ldots, x_m$. We need

$$\Pr[X_n = x_0, X_{n+1} = x_1, \ldots, X_{n+m} = x_m \mid X_n = x] = \Pr[X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m \mid X_0 = x]$$

□

Recall $\mathcal{F}_T$ where T is a stopping time:

$$A \in \mathcal{F}_T \Leftrightarrow A \cap \{T \leq n\} \in \mathcal{F}_n$$

Proof. Key fact: $\{T = n\} \in \mathcal{F}_n$. Given $\{T = n\}$ and the $\mathcal{F}_T$ info, you basically know some $\mathcal{F}_n$ fact. □